

Probleme der kulturvergleichenden Umfrageforschung

Haas, Helene

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Haas, H. (2007). Probleme der kulturvergleichenden Umfrageforschung. *interculture journal: Online-Zeitschrift für interkulturelle Studien*, 6(5), 3-20. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-451321>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Probleme der kulturvergleichenden Umfrageforschung

Helene Haas

Diplom-Kulturwirtin, Doktorandin
an der Universität Passau

Abstract

This paper calls attention to important methodological problems in cross-cultural survey research and questionnaire construction. Different sorts of bias limit the comparability of data: respondents give socially desirable rather than true answers, their cultural response-sets show different tendencies toward acquiescence and extreme-response checking, they have different denotations and connotations of key expressions in mind and find some items ambiguous or perceive them as double-barrel questions. The most severe limitations to comparative validity lie in the difficulty of finding equivalent translations for questionnaires. Scalar equivalence is related both to translation and cultural response-sets. Adequate sampling is an additional challenge to comparative cross-cultural studies. Cross-cultural survey researchers recommend a wide range of validation procedures, but a unified and secured methodology has yet to be developed. Until that time, statements about cultural differences should not be accepted uncritically by researchers or the public.

1. Methodische Probleme beim Kulturvergleich

Kulturvergleichende Studien erfreuen sich bereits seit Jahrzehnten stetig wachsender Beliebtheit. Das öffentliche Interesse ist groß, entsprechend stark stieg die Forschungsaktivität in diesem Bereich an (Berry et al. 2002:6f).

Wie aber steht es mit der Validität der wissenschaftlich ermittelten Kulturbeschreibungen? Vielfach ist nicht bekannt, dass sich der Forschungsbereich bislang noch auf kein ausgereiftes Instrumentarium zur Untersuchung kulturbedingter Unterschiede stützen kann. Kaase (1999:59) bezeichnet die international vergleichende Sozialforschung als „keineswegs befriedigend“. Das Desiderat eines gesicherten methodischen Regelwerks für den Kulturvergleich besteht schon lange (Berry 1980:1), doch die Debatte zur seiner Entwicklung kommt zu keinem Ergebnis.

Je nach fachlicher Provenienz untersuchen kulturvergleichende Forscher ihren Gegenstand mit unterschiedlichen Instrumenten. Von allen Methoden kommt – natürlich aus Gründen der Praktikabilität – die Umfrage am häufigsten zum Einsatz (Porst 2000:16). Dieser Beitrag konzentriert sich daher auf die Problemfelder der kulturvergleichenden Umfrageforschung im Bereich der Fragebogenentwicklung und der Auswahl von Testpersonen. Wichtig sind hier der Einfluss sozialer Erwünschtheit auf das Antwortverhalten der Befragten, kulturbedingt unterschiedliche Antwortstile, sprachliche Äquivalenzprobleme bei der Übersetzung und Skalenerstellung, divergierende Interpre-

tationen von Schlüsselbegriffen sowie unklare oder mehrdimensionale Formulierungen. Auch die Form der Stichprobe beeinflusst die Qualität einer kulturvergleichenden Umfrage.

2. Soziale Erwünschtheit

Eine Befragung sammelt Daten in Form von Meinungen der Probanden. Damit stützt sich die Umfrageforschung auf Aussagen, die zum Großteil nicht objektiv überprüfbar sind. Testpersonen neigen dazu, von der Wahrheit abweichende Angaben zu machen, um sich in besseres Licht zu rücken. Viele von ihnen wollen vermeintlichen Erwartungshaltungen entsprechen und negative Reaktionen auf ihr Verhalten von Seiten des Interviewers minimieren (Diekmann 2004:383). Aus diesem Grund machen sie verstärkt Aussagen, die anerkannten Meinungen entsprechen, während abweichende Einstellungen eher verschwiegen werden (Smith 2004b:439). Von der Außenwirkung abgesehen, fällt das im Bewusstsein gespeicherte Selbstbild des Einzelnen wahrscheinlich auch positiver aus als das Fremdbild, das andere sich von ihm machen.

Dieses Phänomen wird im Deutschen als soziale Erwünschtheit bezeichnet, in englischsprachigen Publikationen ist von *social desirability* die Rede. Scholl (1993:70) spricht von einer „Strategie der Selbstpräsentation“, Hui (2001:23) von „Eindrucksmanagement“. Shavitt et al. (2004:11f) unterscheiden zwischen absichtlichem Eindrucksmanagement (*impression management*), also der bewussten Korrektur von Aussagen zur Erhöhung des eigenen sozialen Ansehens, und einer selbsttäuschenden Verbesserung (*self-deceptive enhancement*). Hier glaubt der Befragte selbst an die überhöhte Darstellung der eigenen Person.

Soziale Erwünschtheit tritt immer dann auf, wenn in Umfragen sensible Themen angesprochen werden. Vor allem aber in der kulturvergleichenden Fragebogenforschung stellt das Phänomen eine schwerwiegende Fehlerquelle dar (Johnson / van de Vijver 2003:197), da die systematische Ergebnisverzerrung durch soziale Erwünschtheit in einzelnen Kulturkreisen unterschiedlich ausfällt. Aufgrund unterschiedlicher Wertorientierungen haben die Probanden variierende Idealvorstellungen menschlichen Verhaltens und „korrigieren“ ihre Aussagen folglich auch in verschiedene Richtungen (Scholl 1993:67, Smith 2004b:439). Nicht nur die Richtung, sondern vor allem auch die Stärke der Modifikation von Antworten variiert kulturbedingt, da bestimmte Themen in den einzelnen Kulturen als unterschiedlich sensibel eingestuft werden. Während ein Begriff in einer Gesellschaft sehr emotional belegt ist, kann er anderswo als eher neutraler oder technischer Ausdruck verstanden werden (Scheuch 1968:182). Der Zusammenhang mit dem Phäno-

men sozialer Erwünschtheit erschließt sich intuitiv: Die Tendenz, Antworten im Hinblick auf ihre soziale Akzeptanz zu editieren, wirkt sich umso stärker aus, je sensibler der Bereich für den Befragten ist. Die distanzierte und entspannte Betrachtung eines Themas erleichtert dem Probanden eine wahrheitsgemäße Antwort, während er sich bei Fragen über emotional stark aufgeladene Bereiche oder Tabuthemen unwohl fühlt und meint, sein Gesicht wahren zu müssen.

Johnson und van de Vijver (2003:201) glauben, ein konsistentes Muster kulturbedingter Unterschiede in der Tendenz zu sozialer Erwünschtheit gefunden zu haben. Aus der Betrachtung aktueller Studien ziehen sie den Schluss, Testpersonen aus schwächeren sozialen Gruppen seien aufgrund ihres geringeren Selbstvertrauens in höherem Maße bestrebt, einen guten Eindruck zu machen, während Mitglieder selbstbewusst auftretender, starker sozialer Gruppen eher den Mut zur Wahrheit hätten. In reicheren Ländern sei die Tendenz zu sozialer Erwünschtheit weniger stark ausgeprägt als in ärmeren Ländern. Auch wer soziale Macht oder einen höheren Bildungsstand habe, empfinde ein schwächeres Bedürfnis, sich in besseres Licht zu rücken. Dieses Modell von Johnson und van de Vijver ist jedoch umstritten. Der Einfluss kultureller Faktoren auf die Tendenz zu sozialer Erwünschtheit ist noch nicht zur Genüge erforscht, doch die bisherigen Befunde stützen die Annahme, dass sich das Phänomen asymmetrisch auf die Ergebnisse kulturvergleichender Fragebögen auswirkt.

Vielfältige Vorschläge (Fisher 1993:303f, Holtgraves / Eck / Lasky 1997, Hui 2001:21ff) zur Minimierung dieses Verzerrungseffektes zielen darauf ab, den Befragten das Gefühl des sozialen Drucks zu nehmen. Probanden müssen auch bei einer ehrlichen Antwort noch ihr Gesicht wahren können. Anonymitätszusicherungen gelten in der allgemeinen Umfragepraxis daher meist als selbstverständlich. Trotz aller Bemühungen bleiben Effekte der sozialen Erwünschtheit jedoch weiterhin bestehen und relativieren die Vergleichbarkeit der Ergebnisse kulturvergleichender Studien.

3. Unterschiedliche Antwortstile

Eng verbunden mit der Tendenz zu sozialer Erwünschtheit ist das Problemfeld kulturell variierender Antwortstile (Diekmann 2004:382, van Herk / Poortinga / Verhallen 2004:347). Probanden aus verschiedenen Kulturkreisen reagieren auf den gleichen Fragebogen in unterschiedlicher Weise (Porst 2000:47). Vor allem die Neigung zur sogenannten Acquieszenz und die Tendenz, auf einer Antwortskala extreme Optionen zu wählen, unterscheiden sich.

Acquieszenz bezeichnet die Neigung vieler Testpersonen, unabhängig von ihren wahren Einstellungen Fragen affirmativ zu beantworten. An Stelle der notdürftigen Übersetzung des englischen Begriffs *acquiescence* ist im Deutschen auch gelegentlich die Rede von „inhaltsunabhängiger Zustimmungstendenz“. Hofstede (1980:56) verwendet die plakative Bezeichnung „yes-man-ship“, Smith (2004b:440) spricht von „yea-saying“.

Laut Scholl (1993:66) sind für diesen Effekt ähnliche Faktoren verantwortlich wie für die Neigung zu sozial erwünschten Antworten, da beide Phänomene in dem Bestreben des Probanden wurzeln, beim Interviewer einen guten Eindruck zu hinterlassen. Demnach weisen unsichere Testpersonen aus der Unterschicht oder mit niedrigem Bildungsstand eine höhere inhaltsunabhängige Zustimmungstendenz auf, da sie sich der Befragungssituation häufig kognitiv nicht gewachsen fühlen und Missbilligung von Seiten des Interviewers vermeiden möchte. Ein gänzlich anderer Grund für Acquieszenz kann jedoch auch in niedriger Motivation des Befragten liegen: Anstatt sich mit dem Inhalt der Fragen kognitiv auseinander zu setzen, antwortet dieser unreflektiert affirmativ, um das Interview auf dem Wege des geringsten Widerstandes hinter sich zu bringen (Smith 2004a:54, van de Vijver / Ploubidis / van Hemert 2004:7).

Kulturvergleichende Studien müssen sich der Erkenntnis stellen, dass die Neigung zur Acquieszenz systematisch zwischen Gesellschaften variiert. So stellen Diekmann (2004:386) und Smith (2004a:55) diese Tendenz verstärkt in nicht-westlichen Kulturkreisen fest, während van Herk, Poortinga und Verhallen (2004:357) auch zwischen europäischen Ländern deutliche Unterschiede konstatieren. Ihren Studien zufolge tendieren Griechen innerhalb der europäischen Union am stärksten zu Acquieszenz, doch auch Spanier und Italiener erreichen höhere Werte als beispielsweise Briten, Deutsche und Franzosen.

Um den Grad der Antwortverzerrung durch Acquieszenz zu bestimmen empfehlen Umfrageforscher in Bezug auf einen Themenbereich sowohl positiv als auch negativ formulierte Fragen zu stellen (Welkenhuysen-Gybels / Billiet / Cambré 2003:711, Smith 2004a:51, Smith 2004b:440, van de Vijver / Ploubidis / van Hemert 2004:17, van Herk / Poortinga / Verhallen 2004:348). Durch die gegensätzliche Polung widersprechen sich zustimmende Antworten und deuten damit auf einen hohen Acquieszenzfaktor hin (Welkenhuysen-Gybels / Billiet / Cambré 2003:706). Obwohl diese Vorgehensweise einhellig empfohlen wird, kommt sie in der Praxis nur sehr selten zur Anwendung (Smith 2004a:51).

Der zweite wichtige Problemkreis formaler Antwortstile ist die Tendenz zur Wahl extremer Skalenpunkte. Manche Testpersonen wählen überwiegend Antwortoptionen an den Endpunkten

der jeweils angebotenen Skalen, während sich andere Probanden grundsätzlich eher für Mittelpositionen entscheiden. In der englischen Literatur ist hier von *extreme response checking* die Rede, im Deutschen findet sich gelegentlich die Bezeichnung „Extremwert-Antwort“ (Scholl 1993:62). US-amerikanischen Studien zufolge hängt dieses Phänomen von der ethnischen Zugehörigkeit der Befragten ab. Demnach wählen schwarze und hispanische amerikanische Studenten häufiger extreme Skalenwerte als ihre weißen, nicht-hispanischen Kommilitonen. Auch Untersuchungen formaler Antwortstile in verschiedenen Kulturkreisen offenbarten deutliche Unterschiede. Cheung und Rensvold (2000:189) sehen die Ursache dafür in verschiedenen Wertorientierungen: Während Mitglieder von Kulturen mit starker Tendenz zu Extremwert-Antworten durch deutliche Aussagen Ehrlichkeit und Überzeugung demonstrieren wollen, möchten Angehörige von Kulturen mit Hang zu Mittelwerten eher bescheiden erscheinen.

Sind demographische oder schichtenspezifische Faktoren wie Bildungsniveau, Alter oder Einkommen nicht auch ein Grund für Unterschiede in formalen Antwortstilen? Sicherlich, doch der kulturelle Hintergrund der Probanden spielt weiterhin eine wichtige Rolle. Er verzerrt die Umfrageergebnisse und macht damit eine einfache Datenauswertung durch das Errechnen von Mittelwerten unmöglich (Cheung / Rensvold 2000:189). Da jedoch noch kein ausgereifter Korrekturmechanismus entwickelt werden konnte und eine Analyse formaler Antwortstile äußerst aufwändig ist, beschäftigt sich die Mehrzahl der kulturvergleichenden Studien in der Praxis nicht mit diesem Phänomen. Van Herk, Poortinga und Verhallen (2004:358) warnen also zu Recht vor einer Vernachlässigung dieser Effekte im Kulturvergleich.

4. Verständnis von Schlüsselbegriffen

In den einzelnen Kulturkreisen aktivieren bestimmte Schlüsselbegriffe überraschend unterschiedliche Vorstellungsinhalte. Testpersonen aus verschiedenen Kulturen legen abweichende Vergleichsmaßstäbe an die Bewertung der abgefragten Sachverhalte an (Sudman / Bradburn / Schwarz 1996:251, Braun / Scott 1998:143). Heine demonstriert dies an einem plakativen Beispiel: So findet die Aussage „I am short“ je nach durchschnittlicher Körpergröße in der jeweiligen Kultur unterschiedlich starke Zustimmung (Heine 2004:4f). Ein 1,75m großer Mann wird diese Aussage in einer Region mit niedriger durchschnittlicher Körpergröße sicherlich ablehnen. Lebt er jedoch unter sehr hoch gewachsenen Mitmenschen, tendiert er stärker zur Zustimmung. Hängt die Beantwortung derartiger Einschät-

zungsfragen von kulturell verankerten Vergleichsmaßstäben ab, so kann man die Befragungsergebnisse aus einzelnen Ländern nicht miteinander vergleichen.

Diese Effekte zu korrigieren ist äußerst schwierig, da sich dem Fragebogenforscher nur sehr schwer erschließt, wann kulturbedingt unterschiedliche Vergleichsmaßstäbe angelegt werden. Laut Harkness, van de Vijver und Mohler (2003) tendieren die Entwickler von Fragebögen daher meist dazu, über Kulturgrenzen hinweg einheitliche Bedeutungen von Begriffen anzunehmen, ohne diese Prämissen zu überprüfen. In der Regel sind sich Forscher nicht bewusst, wie die Schlüsselbegriffe ihrer Fragen in fremden Kulturen verstanden werden. Diese Sorglosigkeit öffnet die Tür für potenzielle Fehlerquellen.

5. Unklarheit und Mehrdimensionalität

Eines der häufigsten Probleme von Umfragen erscheint banal, hat aber schwerwiegende Folgen: Die Testperson versteht die Frage nicht. Fragebögen erlauben kein Nachhaken bei Unklarheiten, keine Klarstellungen missverständlicher Formulierungen und keine Rückfragen. Vor allem bei einer schriftlichen Befragung ohne persönlichen Kontakt zum Interviewer verfügt der Proband also über keinerlei Hilfsmittel zur Interpretation. Wie eine derart auf sich gestellte Testperson die Bedeutung unklarer Fragen zu erschließen versucht, erforschte Braun (2003:59) unter dem Blickwinkel der Kognitionspsychologie. Probanden suchen demnach die Schuld bei sich selbst und bemühen sich, die Intention des Forschers durch eigene Interpretationen zu erschließen. Dabei konzentrieren sie sich nicht nur auf die wörtliche Formulierung der vorliegenden Frage, sondern ziehen auch alle Arten von Kontextinformationen als Interpretationshilfen heran (Sudman / Bradburn / Schwarz 1996:267). Dies kann vorhergehende Fragen mit einschließen, sich jedoch auch auf die Interviewsituation oder sogar die momentane Befindlichkeit des Einzelnen und seine aktuellen Erfahrungen beziehen. Wenn die Antworten derart kontextabhängig sind, können die Umfrageergebnisse nicht mehr verglichen werden.

Dieser Effekt ist bereits im monokulturellen Umfeld schwer zu kontrollieren, potenziert sich jedoch bei kulturvergleichenden Studien (Smith 2003:84). Der zur Interpretation einer Frage relevante Kontext umfasst hier nicht mehr nur den Rest des Fragebogens und persönliche Erfahrungen des Probanden, sondern schließt auch die Standardisierungen der jeweiligen Kultur mit ein (Braun 2003:60). Befragte aus verschiedenen Ländern können bei Unklarheiten daher ganz unterschiedliche Kontextinformationen heranziehen und interpretieren die Frage folglich kulturbedingt auf unterschiedliche Weise. Vergleichbar-

keit der Umfrageergebnisse ist damit nicht mehr gegeben. Unklar erscheint eine Frage vor allem dort häufig, wo ihr Thema weniger stark im öffentlichen Bewusstsein steht als dies in der Kultur des Forschers der Fall ist. In diesem Fall können Testpersonen die für ein korrektes Verständnis der Frage notwendigen Hintergrundinformationen nicht so leicht abrufen, wie der Entwickler des Fragebogens dies vermutet hatte.

Neben Unklarheiten können auch so genannte mehrdimensionale Fragen zu Missverständnissen führen. Von Mehrdimensionalität spricht man, wenn in einer Frage zwei oder mehr verschiedene Sachverhalte gleichzeitig angesprochen werden und eine einheitliche Antwort dafür verlangt ist. Ein hypothetisches Beispiel dafür wäre die Frage „Verbringen Sie viel Zeit mit Ihrer Familie und Ihren Freunden?“ Wie geht ein Single, der seine ganze Freizeit im Freundeskreis verbringt, mit dieser Frage um? Der eine wird den Mittelpunkt der Skala wählen, weil er beide Aspekte berücksichtigt, der andere dagegen sich mit Blick auf die Freunde für einen Extremwert entscheiden. Somit haben die Ergebnisse keinerlei Aussagekraft. Zu derart fehlerbehafteten *double-barrel questions* (Smith 2004b:445, Sudman / Bradburn / Wansink 2004:325) kommt es vor allem dann, wenn der Forscher zwei Dinge als untrennbar verbunden sieht, wo der Befragte eine Unterscheidung vornimmt. Im kulturübergreifenden Rahmen ist dies besonders häufig der Fall, da Wissenschaftler und Zielpersonen oft kulturbedingt unterschiedliche Vorstellungen von den betrachteten Konstrukten haben.

Um Quellen von Unklarheiten oder Mehrdimensionalitäten zu identifizieren, empfiehlt Porst (2000:65f) umfangreiche Vorstudien zur Überprüfung der Funktionsfähigkeit eines Fragebogens. Im kulturvergleichenden Kontext müssen Verständnisproben durchgeführt werden, in deren Rahmen Testpersonen bestimmte Schlüsselbegriffe paraphrasieren und damit ihre Vorstellungsinhalte konkretisieren. Vorstudien dieser Art werden jedoch in der Praxis nur unsystematisch durchgeführt (Fowler 1992:219). Zudem sind sie bezüglich ihrer Anwendung im Kulturvergleich noch nicht methodisch überprüft (Smith 2004b:450) und dürfen daher nicht als Garant für eine perfekte Vergleichbarkeit von Fragebögen verstanden werden. Zur Sicherheit sollte jedes Konstrukt mit mehreren Indikatoren getestet werden, die ähnliche Inhalte mit unterschiedlichen Formulierungen abfragen (Smith 2004b:433). Ist eine dieser Fragen missverständlich und liefert irreführende Ergebnisse, so fällt dieser Fehler im Vergleich mit den Resultaten paralleler Indikatoren sofort ins Auge.

6. Übersetzung

Bei internationalen Umfrageprojekten können die Ergebnisse einzelner Länder nur dann miteinander verglichen werden, wenn jede Frage in allen Sprachversionen die gleiche Bedeutung hat. Dies erscheint selbstverständlich, doch in der Übersetzungspraxis erweist es sich häufig als äußerst problematisch, diese Grundvoraussetzung zu gewährleisten (Harkness 2003:36, Smith 2004b:432). In jedem Ausgangsfragebogen decken einige Wörter und Ausdrücke ein derart vielschichtiges Bedeutungsspektrum ab, dass in der Zielsprache keine vollkommen äquivalente Formulierung zu finden ist. Dies betrifft nicht nur die Ebene der Denotationen: Gerade im Bereich der Konnotationen können schon kleinste Variationen die Bedeutung einer Frage stark verändern (Braun / Scott 1998:129). Kann ein Übersetzer Schlüsselbegriffe einer Frage nicht mit ihren vollständigen Denotationen und Konnotationen in eine andere Sprache übertragen, so muss er sich auf einen Begriff in der Zielsprache festlegen, dessen Aussagegehalt nicht in allen Facetten mit dem der Originalversion deckungsgleich ist. Im Extremfall resultiert aus dem Übersetzungsvorgang eine völlig andere Frage (Harkness 2003:49, Harkness 2004:456).

Besonders gravierend wirkt sich die Problematik aus, wenn ein ganzes Konzept in der Ausgangssprache durch einen Einzelbegriff prägnant beschrieben ist, während es in der Zielsprache keinen Ausdruck dafür gibt. Dies ist der Fall, wenn das betreffende Phänomen kulturbedingt unterschiedlich wahrgenommen wird oder in der Zielkultur im Extremfall nicht einmal existiert (Scheuch 1968:182). Ein für westliche Wissenschaftler überraschendes Beispiel findet sich bei Frey (1963), demzufolge Bewohner ländlicher Gegenden in der Türkei kein Wort für „Problem“ oder „Loyalität“ kennen.

Aber auch Unterschiede in der Konstruktion einzelner Sprachen stellen Übersetzer vor schwierige Entscheidungen. So kann man im Englischen das Geschlecht einer Person leicht unidentifiziert lassen, während im Spanischen, Französischen oder auch im Deutschen zwischen maskulinen und femininen Formen unterschieden wird. Beinhaltet der englische Ausgangsfragebogen beispielsweise die Formulierung „Now think about your best friend...“, so schließt dies sowohl weibliche als auch männliche Freunde ein. Soll die Frage in der deutschen Version keine eingeschränkte Bedeutung bekommen, müssen sowohl die maskuline als auch die feminine Form des Wortes genannt werden: „Und nun zu Ihrem besten Freund/Ihrer besten Freundin...“ (Harkness 2003:50, Harkness 2004:458f). Auch die Anrede der Testperson kann sich gelegentlich als problematisch erweisen, da in Sprachen wie dem Deutschen, Französischen oder Spanischen zwischen der informellen Anrede du/tu/tú und der Höf-

lichkeitsform Sie/Vous/Usted unterschieden wird, während diese Unterscheidung im Englischen nicht existiert. Der zumeist in englischer Sprache abgefasste Originalfragebogen gibt keine Anhaltspunkte, welche Form in den Übersetzungen zu wählen ist. Werden nur Erwachsene befragt, lässt sich dieses Problem leicht mit der Verwendung der Höflichkeitsform lösen. Beinhaltet die Stichprobe jedoch auch Kinder, so besteht die Gefahr, diese durch Verwendung der für sie ungewohnten Höflichkeitsform zu befremden. Diffizil ist auch die Frage, wie Jugendliche angesprochen werden möchten (Harkness 2004:459f).

Angesichts der vielfältigen Fallstricke überprüfen inzwischen fast alle kulturvergleichenden Umfrageprojekte ihre übersetzten Fragebögen. Meistens werden die Versionen aus der Zielsprache von einem unabhängigen Übersetzer in die Ausgangssprache zurück übertragen und mit der Originalversion verglichen (Smith 2004b:447). Diese einfache Rückübersetzung kann jedoch keinesfalls alle Übersetzungsprobleme eliminieren (Harkness / Schoua-Glusberg 1998:111f, Harkness 2003:42), weshalb Johnson (1998:18) in diesem Zusammenhang vor einem „falschen Gefühl der Sicherheit“ warnt. Um das Äquivalenzproblem bei Übersetzungen zu minimieren suchen Umfrageforscher daher zusätzlich nach einfachen Grundregeln zur Fragebogengestaltung, die leichtere Übersetzbarkeit gewährleisten sollen. Brislin empfiehlt beispielsweise (1986:144ff), kurze Fragen in einfachem Vokabular zu stellen und Metaphern, umgangssprachliche Ausdrücke sowie vage Begriffe zu vermeiden. Dies erscheint vernünftig, jedoch reichen auch diese Rezepte nicht aus, um das Verzerrungspotenzial von Übersetzungen vollständig zu beseitigen.

So viele Empfehlungen die Forschung für eine verbesserte Übersetzungsmethodik auch gibt, so häufig wird die Problematik in der Praxis vernachlässigt. Ein berühmtes Beispiel sind die groß angelegten Studien von Geert Hofstede zu seinem wegweisenden Werk „Culture's Consequences“. Hier wurden die meisten Sprachversionen der Fragebögen nicht einmal durch Rückübersetzungen kontrolliert, sondern lediglich zweisprachigen Mitarbeitern von IBM zur Durchsicht vorgelegt. Hofstede rechtfertigt diese bequeme Vorgehensweise mit dem ohnehin immensen organisatorischen Aufwand und den finanziellen und zeitlichen Restriktionen der Forschung (Hofstede 1980:45ff). Diesen unterliegen auch aktuelle kulturvergleichende Projekte, doch im Lichte der neuen Erkenntnisse auf dem Gebiet der Fragebogenforschung könnte eine ambitionierte Studie das Übersetzungsproblem heute nicht mehr auf diese Weise abhandeln.

Da die vielfältigen Verzerrungseffekte durch Übersetzung von Fragebögen nicht genau bestimmt werden können (Grimm et al. 1999:490) ist es unerlässlich, detaillierte Informationen über

den Übersetzungsvorgang zu speichern (Brislin 1986:152, Harkness 2003:43, Harkness 2004:470ff). Nur wenn konkrete Problempunkte, alternative Möglichkeiten für strittige Formulierungen, die Ergebnisse von Qualitätsprüfungen und daraufhin vorgenommene Änderungen aufgezeichnet werden, kann die Äquivalenz auch später noch evaluiert werden. Diese Dokumentation ist laut Harkness (1999:134) bei den meisten Studien jedoch höchstens rudimentär vorhanden.

7. Skalenäquivalenz

Ein zentraler Punkt bei der Konzeption jedes Umfrageprojekts ist die Entscheidung, in welcher Form die Daten erhoben werden sollen. Forscher können wählen zwischen offenen Fragen, die Raum zur Erschließung neuer Aspekte lassen (Porst 2000:48, Diekmann 2004:409), und geschlossenen Fragen, die Probanden eine Antwortskala anbieten und die statistische Auswertung der Antworten erlauben. Da nur letztere Form einen direkten Vergleich der Ergebnisse einzelner Testpersonen oder zwischen Gruppen von Probanden (Diekmann 2004:408) zulässt, kommen in der kulturvergleichenden Umfrageforschung fast ausschließlich geschlossene Fragen zum Einsatz. Wichtig ist hier, dass auf den Antwortskalen in allen Sprachvarianten gleichbedeutende Alternativen angeboten werden. Nur so können die in den einzelnen Kulturen gewonnenen Ergebnisse miteinander verglichen werden.

Antwortskalen können höchst unterschiedlich gestaltet werden. So haben Forscher die Wahl zwischen numerischen Skalen mit bloßer Benennung der Endpunkte und verbalisierten Skalen mit Beschreibung jeder Antwortoption. Bezüglich numerischer Skalen stellt sich weiterhin die Frage, ob eine ungerade Anzahl von Skalenpunkten und damit eine formale Mittelposition angeboten werden soll oder ob eine gerade Anzahl von Punkten zu bevorzugen ist. In der Fragebogenforschung gibt es dazu keine vorherrschende Meinung. Während eine formale Mittelposition oft von unentschiedenen Testpersonen als bequeme Ausweichmöglichkeit benutzt wird und damit die Aussagekraft der Antworten in Frage stellen kann, nimmt eine gerade Skala den Befragten die Möglichkeit, sich bewusst neutral einzustufen. Auch über die optimale Anzahl von Skalenpunkten konnte in der Forschung noch keine Einigkeit erzielt werden (Diekmann 2004:329).

Im Rahmen kulturvergleichender Umfragen ist zu beachten, dass die Verwendung des gleichen Skalentyps in verschiedenen Kulturen und Sprachen nicht notwendigerweise Bedeutungsäquivalenz gewährleistet. Zurückzuführen ist dies auf die kulturbedingt unterschiedliche Tendenz zur Extremwert-Antwort,

die bereits in Abschnitt 3 besprochenen wurde. Basierend auf der Beobachtung, dass Asiaten eher zur Wahl moderater Skalenpunkte nahe der Mitte tendieren, halten Harkness, van de Vijver und Johnson (2003:29) beispielsweise eine bipolare Fünfpunktskala mit starken Endbezeichnungen in Japan für weniger gut geeignet als eine längere Skala. In einer Kultur mit starker Neigung zur Extremwert-Antwort könnte die Fünfpunktskala dagegen durchaus adäquat sein. Auch kennen manche Kulturen Unglückszahlen, was sich bei der Verwendung numerischer Skalen bemerkbar machen kann (Mohler / Smith / Harkness 1998:167, Harkness 2003:53, Smith 2004b:436). Im kulturvergleichenden Kontext erweisen sich numerische Skalen dennoch meist als praktikabler als verbalisierte, da es bei der Konstruktion letzterer äußerst schwierig sein kann, in mehreren Sprachen äquivalente Benennungen für die Antwortoptionen zu finden. Das bereits besprochene Übersetzungsproblem hat hier seine gravierendsten Auswirkungen: Bereits die kleinste Änderung in den Bedeutungsnuancen der Antwortoptionen kann Probanden zur Wahl eines anderen Skalenpunktes bewegen (Mohler / Smith / Harkness 1998:164). In allen Sprachen vollständig bedeutungsäquivalente Benennungen der Skalenpunkte zu finden ist laut Harkness (2004:458) jedoch unmöglich, da der Forscher dazu die Intensität einzelner Ausdrücke in den verschiedenen Sprachen kennen müssten. Werden beispielsweise die deutschen Begriffe „stimme völlig/eher/teilweise zu“ beziehungsweise „lehne teilweise/eher/völlig ab“ als genauso stark empfunden wie die englischen Ausdrücke „agree completely/tend to agree/somewhat agree“ beziehungsweise „somewhat disagree/tend to disagree/completely disagree“? Smith (2004c) führte dazu umfangreiche und methodisch sehr anspruchsvolle Studien mit Deutschen und Amerikanern durch und stellte fest, dass vielen Basisbegriffen von Antwortskalen, beispielsweise agree/stimme zu und disagree/lehne ab oder important/wichtig und unimportant/unwichtig, im Deutschen höhere Intensitäten zugeschrieben werden als im Englischen. Anders verhält es sich bei Modifikatoren. So ist beispielsweise das englische „definitely“ ein stärkerer Ausdruck als der deutsche Modifikator „bestimmt“, weshalb die allgemein übliche Übersetzung des englischen „definitely“ mit diesem Begriff problematisch ist. Auch „a lot“ mit „ziemlich“ zu übersetzen ist nicht zu empfehlen, da der englische Ausdruck verstärkend wirkt, das deutsche Wort die Intensität einer Aussage dagegen eher abschwächt (Smith 2004c:9f). Diese Beispiele zeigen das beträchtliche Verzerrungspotenzial, das Übersetzungen verbalisierter Antwortskalen mit sich bringen. Die vorherrschende Meinung empfiehlt daher für kulturvergleichende Projekte die Benutzung numerischer Skalen. (Johnson 1998:16f, Smith 2004b:435, Sudman / Bradburn / Wansink 2004:329).

8. Stichproben

Die Auswahl repräsentativer Stichproben ist eine allgemeine Problematik der Umfrageforschung, gestaltet sich aber vor allem im interkulturellen Kontext schwierig (Berry / Lonner 1986:85). Defizite in diesem Bereich lassen die Validität vieler kulturvergleichender Studien fraglich erscheinen. Die Qualität einer Stichprobe wird daran gemessen, ob die in ihrem beschränkten Rahmen gewonnenen Erkenntnisse auf die Gesamtpopulation übertragen werden können. Dazu muss die Stichprobe ein möglichst genaues Abbild der Grundgesamtheit, aus der sie gezogen wurde, darstellen.

Der herrschenden Meinung zufolge erlauben nur Zufallsstichproben zuverlässige Rückschlüsse auf eine Grundgesamtheit (Lazerwitz 1968, Berry / Lonner 1986, Harkness 1999, Häder / Gabler 2003, Diekmann 2004). Hier hat jede Person in der betrachteten Grundgesamtheit eine gleich hohe Chance, in die Stichprobe aufgenommen zu werden (Lazerwitz 1968, Friedrichs 1990, Schnell et al. 1995, Diekmann 2004). Im kulturvergleichenden Kontext kann diese aufwändige Methode jedoch oft nicht verfolgt werden, da sie den Zugang zu einem Bevölkerungsregister voraussetzt. Vor allem in Entwicklungsländern ist ein solches häufig nicht vorhanden oder wird den Forschern nicht zugänglich gemacht. Kulturvergleiche stützen sich daher meist auf Quotenstichproben. Bei dieser Technik ist darauf zu achten, dass die Größe demographischer Subgruppen in Relation zur Gesamtbevölkerung ihren Anteilen in der Stichprobe entspricht. In der kulturvergleichenden Forschung besteht weitgehende Einigkeit, dass Zufallsstichproben die wünschenswerte Ideallösung darstellen (Porst 2000:60, Häder / Gabler 2003:121), Quotenstichproben jedoch deutlich praktikabler sind. Vor 20 Jahren beklagten Berry und Lonner (1986:85) den Mangel an spezifischen Techniken der Stichprobenziehung für die kulturvergleichende Fragebogenforschung. Obwohl die Notwendigkeit einer eigens entwickelten, praktikablen Methodik bereits seit Langem anerkannt ist, konnten auf diesem Gebiet noch keine nennenswerten Fortschritte erzielt werden (Häder / Gabler 2003:124).

Neben methodischen Problemen beeinträchtigen im kulturvergleichenden Kontext jedoch noch weitere Einflussfaktoren die Qualität der Stichproben. So weigern sich in verschiedenen Kulturen oft unterschiedlich viele Probanden, an der Befragung teilzunehmen und verändern damit die ursprünglich repräsentativen Stichproben. Je sensibler das Thema einer Befragung in einer Zielkultur ist, desto weniger Testpersonen sind bereit, sich zu äußern (Smith 2004b:441). Verschieden hohe Verweigerungsraten resultieren aber auch aus unterschiedlichen Einstellungen der Bevölkerung gegenüber Meinungsumfragen. So

führt in den USA der Ärger über häufige Störungen durch aufdringliches Telemarketing dazu, dass immer mehr Probanden auch seriöse Umfragen ablehnen (Couper / de Leeuw 2003:158). Eine praktikable Technik zur Korrektur unterschiedlicher Verweigerungsraten konnte bislang noch nicht gefunden werden (Scholl 1993, Diekmann 2004:359).

Aufgrund zeitlicher und finanzieller Restriktionen ist es bei den meisten kulturvergleichenden Studien nicht möglich, die aufwändigen Prozeduren zur Gewinnung einer optimal repräsentativen Stichprobe zu befolgen. Laut Berry und Lonner (1986:87) sowie Häder und Gabler (2003:133) arbeitet die Mehrheit der kulturvergleichenden Untersuchungen nicht mit fundierten Zufallsstichproben, sondern rekrutieren Probanden nur aus leicht zugänglichen Gesellschaftsschichten, was den Aufwand einer Studie beträchtlich reduziert. Berry und Lonner (1986:87) verwenden in diesem Zusammenhang den Begriff *convenience samples*, der mit „Bequemlichkeitsstichproben“ übersetzt werden könnte. In den meisten Fällen werden dabei Befragungen unter Psychologie- und Soziologiestudenten durchgeführt, da diese im Rahmen universitätsnaher Forschung besonders leicht zu erreichen sind. Auch Beschäftigte eines ausgewählten Betriebs bilden häufig eine Stichprobe, wie es beispielsweise in Hofstedes berühmter IBM-Studie der Fall war. Ob derartige Bequemlichkeitsstichproben wirklich ein Abbild ihrer jeweiligen Kultur darstellen können, ist jedoch äußerst fraglich (Kim et al. 1994:9, Johnson et al. 2000:944, Voronov / Singer 2002:466). Die Gruppe der Studenten beinhaltet größtenteils relativ wohlhabende junge Menschen mit überdurchschnittlichem Bildungsniveau. Zu untersuchen wäre zudem, inwiefern intellektuell geschulte Probanden ein anderes Verständnis von Sprache haben und den Text eines Fragebogens anders interpretieren als Testpersonen mit niedrigerem Bildungsniveau. Möglicherweise gehen die mit der Methodik von Umfragen bereits vertrauten Studenten der Soziologie und Psychologie zudem auf andere Weise an die Fragebögen heran als fachfremde Testpersonen. Arbeitet eine Umfrage mit nicht repräsentativen Bequemlichkeitsstichproben, sind Generalisierungen ihrer Ergebnisse zu Aussagen über die Kultur als Ganzes also problematisch.

Repräsentative Stichproben sind dagegen nach Meinung vieler Forscher für den Kulturvergleich gar nicht erforderlich. Ihrer Ansicht nach erfüllen auch *äquivalente* Stichproben die Anforderungen: Solange die demographischen Merkmale der Testgruppen aller Zielkulturen übereinstimmen, lässt demnach auch eine Bequemlichkeitsstichprobe den Vergleich der Kulturen als Ganzes zu. Aus einer Gegenüberstellung der Angaben deutscher und amerikanischer Studenten könnten also durchaus Rückschlüsse auf die deutsche und amerikanische Kultur gezogen werden, da unterschiedliche Ergebnissen bei übereinstim-

menden Rahmenbedingungen hauptsächlich das Produkt kultureller Differenzen sein müssten.

Welche Position man bezüglich repräsentativer oder äquivalenter Stichproben auch bezieht: Jede kulturvergleichende Studie sollte eine detaillierte Dokumentation ihrer Methodik enthalten (Berry / Lonner 1986:85, Häder / Gabler 2003:132). Können repräsentative Stichproben aus Zeit- und Kostengründen nicht gezogen werden, sollte aus den Forschungsberichten klar hervorgehen, wie die Ergebnisse zustande gekommen sind. Vergleichende Aussagen über Kulturen sollten nur mit Hinweis auf die Restriktionen im Forschungsprozess publiziert werden.

9. Schlussfolgerungen

Jeder der genannten Kritikpunkte ist hinlänglich bekannt. Für diesen Artikel wurden sie lediglich zusammengetragen, doch der so entstehende Gesamteindruck lässt die Problematik in einem anderen Licht erscheinen. Wenn ein wissenschaftliches Verfahren so viele Schwachstellen aufweist, die im Verdacht stehen, die Wirklichkeit zu entstellen, dann kann das nicht auf die leichte Schulter genommen werden. Die Kritikpunkte sind alle bekannt, aber die Forschungspraxis nimmt sie kaum zur Kenntnis. Smith (2004:431) weist daraufhin, dass drei Viertel der Studien im Bereich kulturvergleichenden Psychologie ohne jede selbstkritische Vorüberlegung mit der Arbeit beginnen. Dazu passt, dass viele Veröffentlichungen mit dem Hinweis auf die nicht vertretbaren Kosten auf Validierungskontrollen verzichten. Damit präsentiert sich die kulturvergleichende Psychologie in einer paradoxen Situation: Obwohl die berechtigte Kritik immer lauter wird, findet sie bei Forschungsvorhaben kein Gehör. Vielleicht aber ist diese Paradoxie unvermeidbar. Alle genannten Kritikpunkte sind systemisch, das heißt sie beziehen sich auf notwendige Voraussetzungen der Umfragemethode. Harkness, Mohler und van de Vijver (2003:9) (sowie Harkness 1999:137, van de Vijver 1998:42) sprechen genau dieses Problem an. Sie führen aus, dass aufwändige Verfeinerungen der Umfragetechnik nur geringfügige Verbesserungen bewirken und die Verzerrung der Empirie nicht verhindern.

Wenn man den Blickwinkel erweitert, entdeckt man kritische Ansätze, die nicht nur das methodische Verfahren der Umfrage aufs Korn nehmen, sondern den Kulturenvergleich überhaupt. Nach Hansen (2007:169) stellen die Vergleichspole unzulässige Vereinheitlichungen dar. Kulturen – vor allem Nationalkulturen, um die es bei den Vergleichen geht – seien keine homogenen, sondern höchst heterogene und differente Gebilde. Mit dieser Meinung steht Hansen in der modernen Kulturwissenschaft nicht allein, denn auch Autoren wie Ulrich Beck und Clifford

Geertz lehnen die Vereinheitlichung von Nationalkulturen ab. Pauschal, so Hansen, ließen sich Nationen nicht vergleichen, höchstens in engeren Erscheinungsformen. Es mache keinen Sinn, die Ausländerfeindlichkeit in England und Deutschland zu messen, da diese Einstellung mehr vom Bildungsgrad abhängt als von der Nationalität. Denkbar wäre aber ein Vergleich der Ausländerfeindlichkeit von englischen und deutschen Arbeitern. Auch dieser grundsätzliche Zweifel am Erkenntnisgewinn von Nationenvergleichen wiegt schwer.

Was ist zu tun? Bevor blind weiter geforscht wird, wäre das wissenschaftliche Existenzrecht der kulturvergleichenden Umfrageforschung grundsätzlich zu diskutieren. Dabei könnten zukünftige Projekte dadurch helfen, dass sie jeden Schritt ihres Designs und alle Daten offen legen, damit alles der Kritik zugänglich wird.

Literatur

Berry, J. (1980): Introduction to methodology. In: Berry, J. / Triandis, H. (Hg.): *Handbook of cross-cultural psychology. Methodology*. Boston et al.: Allyn and Bacon, S. 1-28.

Berry, J. / Lonner, W. (1986): Sampling and surveying. In: Lonner, W. / Berry, J. (Hg.): *Field methods in cross-cultural research. Cross-cultural research and methodology series*. Beverly Hills et al.: Sage Publications, S. 85-110.

Berry, J. / Lonner, W. / Leroux, J. (Hrsg.) (1973): *Directory of cross-cultural research and researchers*. Bellingham: Western Washington Univ. Press.

Berry, J. / Poortinga, Y. / Segall, M. / Dasen, P. (2002): *Cross-cultural psychology. Research and applications*. Cambridge: Cambridge University Press.

Braun, M. (2003): Communication and social cognition. In: Harkness, J. / van de Vijver, F. / Mohler, P. (Hg.): *Cross-cultural survey methods*. Hoboken: Wiley, S. 57-68.

Braun, M. / Scott, J. (1998): Multidimensional scaling and equivalence. Is having a job the same as working? In: Harkness, J. (Hg.): *ZUMA-Nachrichten. Spezial Band 3 Cross-cultural survey equivalence*. Mannheim: ZUMA, S. 129-144.

Brislin, R. (1986): The wording and translation of research instruments. In: Lonner, W. / Berry, J. (Hg.): *Field methods in cross-cultural research. Cross-cultural research and methodology series*. Beverly Hills et al.: Sage Publications, S. 137-164.

Cheung, G. / Rensvold, R. (2000): Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modelling. *Journal of Cross-Cultural Psychology* 31(2), S. 187 – 212.

Couper, M. / de Leeuw, E. (2003): Nonresponse in cross-cultural and cross-national surveys. In: Harkness, J. / van de Vijver, F. / Mohler, P. (Hg.): *Cross-cultural survey methods*. Hoboken: Wiley, S. 157-178.

Diekmann, A. (2004): *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbeck bei Hamburg: Rowohlt-Taschenbuch-Verlag.

Fisher, R. (1993): Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research* 20(2), S. 303-315.

Fowler, F. (1992): How unclear terms affect survey data. *Public Opinion Quarterly* 56, S. 218-231.

Friedrichs, J. (1990): *Methoden empirischer Sozialforschung*. Opladen: Westdt. Verlag.

Häder, S. / Gabler, S. (2003): Sampling and estimation. In: Harkness, J. / van de Vijver, F. / Mohler, P. (Hg.): *Cross-cultural survey methods*. Hoboken: Wiley, S. 117-134.

Hansen, K. P. (2007): Kritische Überlegungen zum interkulturellen Paradigma. In: Kuhn, B. / Pitz, M. / Schorr, A. (Hg.): *„Grenzen“ ohne Fächergrenzen*. St. Ingbert: Röhrig Universitätsverlag, S. 149-178.

Harkness, J. (1999): In pursuit of quality. Issues for cross-national survey research. *International Journal of Social Research Methodology* 2(2), S. 125-140.

Harkness, J. (2003): Questionnaire translation. In: Harkness, J. / van de Vijver, F. / Mohler, P. (Hg.): *Cross-cultural survey methods*. Hoboken: Wiley, S. 35-56.

Harkness, J. (2004): *Questions in translation and questions in translation*. Vortragsmaterialien für das Sheth Foundation / Sudman Symposium on Cross-Cultural Survey Research, 30.09. – 02.10.2004, Urbana-Champaign, Online-Dokument: <http://www.srl.uic.edu/shethsudman/presentations/harkness.pdf> [08.05.2007].

Harkness, J. / Schoua-Glusberg, A. (1998): Questionnaires in translation. In: Harkness, J. (Hg.): *ZUMA-Nachrichten. Spezial Band 3 Cross-cultural survey equivalence*. Mannheim: ZUMA, S. 87-126.

Harkness, J. / van de Vijver, F. / Johnson, T. (2003): Questionnaire design in comparative research. In: Harkness, J. / van de Vijver, F. / Mohler, P. (Hg.): *Cross-cultural survey methods*. Hoboken: Wiley, S. 19-34.

Harkness, J. / van de Vijver, F. / Mohler, P. (2003): Comparative Research. In: Harkness, J. / van de Vijver, F. / Mohler, P. (Hg.): *Cross-cultural survey methods*. Hoboken: Wiley, S. 3-18.

Heine, S. (2004): *Considering the validity of cross-cultural comparisons of self-report measures*. Vortragsmaterialien für das Sheth Foundation/Sudman Symposium on Cross-Cultural Survey Research, 30.09. – 02.10.2004, Urbana-Champaign, Online-Dokument: <http://www.srl.uic.edu/shethsudman/presentations/heine.pdf> [08.05.2007].

Hofstede, G. (1980): *Culture's consequences*. Newbury Park: Sage Publications.

Holtgraves, T. / Eck, J. / Laskz, B. (1997): Face management, question wording and social desirability. *Journal of Applied Social Psychology* 27(18), S. 1650-1671.

Hui, H. (2001): Double-rating as a method to encourage candid responses to self-report instruments. *Journal of Applied Social Psychology* 31(1), S. 21-30.

Johnson, T. (1998): Approaches to equivalence in cross-cultural and cross-national survey research. In: Harkness, J. (Hg.): *ZUMA-Nachrichten. Spezial Band 3 Cross-cultural survey equivalence*. Mannheim: ZUMA, S. 1-40.

Johnson, T. et al. (2000): Respondent cultural orientations and survey participation. The effects of individualism and collectivism. In: American Statistical Asso-

ciation (Hg.): *Proceedings of the Survey Research Methods Section*. Online-Dokument:
http://www.amstat.org/sections/srms/Proceedings/papers/2000_161.pdf
[08.05.2007].

Johnson, T. / van de Vijver, F. (2003): Social desirability in cross-cultural research. In: Harkness, J. / van de Vijver, F. / Mohler, P. (Hg.): *Cross-cultural survey methods*. Hoboken: Wiley, S. 195-205.

Kaase, M. (Hrsg.) (1999): *Qualitätskriterien der Umfrageforschung*. Berlin: Akad.-Verl.

Lazewitz, B. (1968): Sampling theory and procedures. In: Blalock, A. (Hg.): *Methodology in social research*. New York: McGraw-Hill, S. 236-277.

Mohler, P. / Smith, T. / Harkness, J. (1998): Respondents' ratings of expressions from response scales. A two-country, two-language investigation on equivalence and translation. In: Harkness, J. (Hg.): *ZUMA-Nachrichten. Spezial Band 3 Cross-cultural survey equivalence*. Mannheim: ZUMA, S. 159-184.

Porst, R. (2000): *Praxis der Umfrageforschung*. Stuttgart et al.: Teubner.

Scheuch, E. (1968): The cross-cultural use of sample surveys. Problems of comparability. In: Rokkan, S. (Hg.): *Comparative research across cultures and nations*. Paris / La Hague: Mouton, S. 176-209.

Schnell, R. / Hill, P. / Esser, E. (1995): *Methoden der empirischen Sozialforschung*. München et al.: Oldenbourg.

Scholl, A. (1993): *Die Befragung als Kommunikationssituation*. Opladen: Westdt. Verlag.

Schwarz, N. (1991): *In welcher Reihenfolge fragen? Kontexteffekte in standardisierten Befragungen*. ZUMA-Arbeitsbericht 91/16, Mannheim: ZUMA.

Shavitt, S. et al. (2004): *The relation between cultural orientation & socially desirable responding*. Vortragsmaterialien für das Sheth Foundation / Sudman Symposium on Cross-Cultural Survey Research, 30.09. – 02.10.2004, Urbana-Champaign. Online-Dokument:
<http://www.srl.uic.edu/shethsudman/presentations/shavittetal.pdf>
[08.05.2007]

Smith, P. (2004a): Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology* 35(1), S. 50-61.

Smith, T. (2003): Developing comparable questions in cross-national surveys. In: Harkness, J. / van de Vijver, F. / Mohler, P. (Hg.): *Cross-cultural survey methods*. Hoboken: Wiley, S. 69-92.

Smith, T. (2004b): Developing and evaluating cross-national survey instruments. In: Presser, S. et al. (Hg.): *Methods for testing and evaluating survey questionnaires*. Hoboken, NJ: Wiley-Interscience, S. 431-452.

Smith, T. (2004c): *Methods for assessing & calibrating response scales across countries & languages*. Vortragsmaterialien für das Sheth Foundation/Sudman Symposium on Cross-Cultural Survey Research, 30.09. – 02.10.2004, Urbana-Champaign, Online-Dokument:
<http://www.srl.uic.edu/shethsudman/presentations/smith.pdf> [08.05.2007].

Sudman, S. / Bradburn, N. / Schwarz, N. (1996): *Thinking about answers. The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Sudman, S. / Bradburn, N. / Wansink, B. (2004): *Asking questions. The definitive guide to questionnaire design for market research, political polls, and social and health questionnaires*. San Francisco: Jossey-Bass.

Van de Vijver, F. (1998): Towards a theory of bias and equivalence. In: Harkness, J. (Hg.): *ZUMA-Nachrichten. Spezial Band 3 Cross-cultural survey equivalence*. Mannheim: ZUMA, S. 41-65.

Van de Vijver, F. / Ploubidis, G. / van Hemert, D. (2004): *Toward an understanding of cross-cultural differences in acquiescence & extremity scoring*. Vortragsmaterialien für das Sheth Foundation/Sudman Symposium on Cross-Cultural Survey Research, 30.09. – 02.10.2004, Urbana-Champaign, Online-Dokument: <http://www.srl.uic.edu/shethsudman/presentations/vandevijver.pdf> [08.05.2007].

Van Herk, H. / Poortinga, Y. / Verhallen, T. (2004): Response styles in rating scales. Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology* 35(3), S. 346-360.

Voronov, M. / Singer, J. (2002): The myth of individualism–collectivism - A critical review. *Journal of Social Psychology* 142(4), S. 461–480.

Welkenhuysen-Gybels, J. / Billiet, J. / Cambré, B. (2003): Adjustment for acquiescence in the assessment of the construct equivalence of Likert-type score items. *Journal of Cross-Cultural Psychology* 34(6), S. 702 - 722.